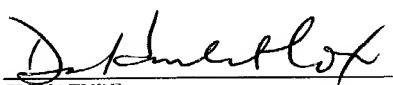


JC04 Rec'd PCT/PTO 0 1 MAY 2001

FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE		ATTORNEY'S DOCKET NUMBER <b>PF-0622 USN</b>
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		U.S. APPLICATION NO. <b>09/831088</b> TO BE ASSIGNED
INTERNATIONAL APPLICATION NO PCT/US99/25820	INTERNATIONAL FILING DATE 03 November 1999	PRIORITY DATE CLAIMED 03 November 1998
TITLE OF INVENTION <b>COENZYME A-UTILIZING ENZYMES</b>		
APPLICANT(S) FOR DO/EO/US <b>INCYTE PHARMACEUTICALS, INC.; TANG, Y. Tom; CORLEY, Neil C.; GUEGLER, Karl J.; GORGONE, Gina A. AZIMZAI, Yalda, KASER, Matthew R.; YUE, Henry</b>		
<p>Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:</p> <ol style="list-style-type: none"> <li><input checked="" type="checkbox"/> This is the <b>FIRST</b> submission of items concerning a filing under 35 U.S.C. 371.</li> <li><input type="checkbox"/> This is a <b>SECOND</b> or <b>SUBSEQUENT</b> submission of items concerning a filing under 35 U.S.C. 371.</li> <li><input type="checkbox"/> This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).</li> <li><input type="checkbox"/> The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).</li> <li><input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2))             <ol style="list-style-type: none"> <li><input type="checkbox"/> is attached hereto (required only if not communicated by the International Bureau)</li> <li><input type="checkbox"/> has been communicated by the International Bureau.</li> <li><input checked="" type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US).</li> </ol> </li> <li><input type="checkbox"/> An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).</li> <li><input checked="" type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))             <ol style="list-style-type: none"> <li><input type="checkbox"/> are attached hereto (required only if not communicated by the International Bureau).</li> <li><input type="checkbox"/> have been communicated by the International Bureau.</li> <li><input type="checkbox"/> have not been made; however, the time limit for making such amendments has NOT expired.</li> <li><input checked="" type="checkbox"/> have not been made and will not be made.</li> </ol> </li> <li><input type="checkbox"/> An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).</li> <li><input checked="" type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).</li> <li><input type="checkbox"/> An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).</li> </ol>		
<b>Items 11 to 16 below concern document(s) or information included:</b>		
<ol style="list-style-type: none"> <li><input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98.</li> <li><input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included.</li> <li><input type="checkbox"/> A <b>FIRST</b> preliminary amendment.             <ul style="list-style-type: none"> <li><input type="checkbox"/> A <b>SECOND</b> or <b>SUBSEQUENT</b> preliminary amendment.</li> </ul> </li> <li><input type="checkbox"/> A substitute specification.</li> <li><input type="checkbox"/> A change of power of attorney and/or address letter.</li> <li><input checked="" type="checkbox"/> Other items or information.</li> </ol> <p>1) Transmittal Letter (2 pp, in duplicate)  2) Return Postcard  3) Express Mail Label No. <b>EL 856 113 172 US</b>  4) Request to Transfer</p>		

U.S. APPLICATION NO. <b>09/831088</b> TO BE ASSIGNED		INTERNATIONAL APPLICATION NO.: PCT/US99/ 25820		ATTORNEY'S DOCKET NUMBER PF-0622 USN	
17. <input checked="" type="checkbox"/> The following fees are submitted: <b>BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5)):</b> Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.....\$1000.00 <input type="checkbox"/> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..\$860.00 International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....\$710.00 <input checked="" type="checkbox"/> <b>International preliminary examination fee paid to USPTO (37 CFR 1.482)</b> <b>but all claims did not satisfy provisions of PCT Article 33(1)-(4)).....\$690.00</b> <input type="checkbox"/> International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4)).....\$100.00					
<b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b>				\$690.00	
Surcharge of <b>\$130.00</b> for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				\$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	20 =	0	X \$ 18.00	\$	
Independent Claims	2 =	0	X \$ 80.00	\$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$270.00	\$	
<b>TOTAL OF ABOVE CALCULATIONS =</b>				\$690.00	
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				\$	
<b>SUBTOTAL =</b>				\$690.00	
Processing fee of <b>\$130.00</b> for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				\$	
<b>TOTAL NATIONAL FEE =</b>				\$690.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). <b>\$40.00</b> per property +				\$	
<b>TOTAL FEES ENCLOSED =</b>				\$690.00	
				Amount to be Refunded	\$
				Charged:	\$
a. <input type="checkbox"/> A check in the amount of \$ _____ to cover the above fees is enclosed. b. <input checked="" type="checkbox"/> Please charge my Deposit Account No. <u>09-0108</u> in the amount of \$ <u>690.00</u> to cover the above fees. c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. <u>09-0108</u> . A duplicate copy of this sheet is enclosed.					
<b>NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must          be filed and granted to restore the application to pending status.</b>					
SEND ALL CORRESPONDENCE TO:  INCYTE GENOMICS, INC. 3160 Porter Drive Palo Alto, CA 94304					
 SIGNATURE					
NAME: Diana Hamlet-Cox					
REGISTRATION NUMBER: 33,302					
DATE: <u>1</u> / <u>May</u> 2001					

## COENZYME A-UTILIZING ENZYMES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of coenzyme A-utilizing enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of neoplastic, immunological, neurological, vesicle trafficking, and muscle disorders.

### BACKGROUND OF THE INVENTION

A large number of cellular biosynthetic intermediary metabolism processes involve intermolecular transfer of carbon atom-containing substrates (carbon substrates). Examples of such reactions include the tricarboxylic acid cycle, synthesis of fatty acids and long-chain phospholipids, synthesis of alcohols and aldehydes, synthesis of intermediary metabolites, and reactions involved in the amino acid degradation pathways. Some of these reactions require input of energy, usually in the form of conversion of ATP to either ADP or AMP and pyrophosphate.

In many cases, a carbon substrate is derived from a small molecule containing at least two carbon atoms. The carbon substrate is often covalently bound to a larger molecule which acts as a carbon substrate carrier molecule within the cell. In the biosynthetic mechanisms described above, the carrier molecule is coenzyme A. Coenzyme A (CoA) is structurally related to derivatives of the nucleotide ADP and consists of 4'-phosphopantetheine linked via a phosphodiester bond to the alpha phosphate group of adenosine 3',5'-bisphosphate. The terminal thiol group of 4'-phosphopantetheine acts as the site for carbon substrate bond formation. The predominant carbon substrates which utilize CoA as a carrier molecule during biosynthesis and intermediary metabolism in the cell are acetyl, succinyl, and propionyl moieties, all of which are examples of acyl groups. Other carbon substrates include enoyl lipid, which acts as a fatty acid oxidation intermediate, and carnitine, which acts as an acetyl-CoA flux regulator/mitochondrial acyl group transfer protein. Acyl-CoA and acetyl-CoA are synthesized in the cell by acyl-CoA synthetase and acetyl-CoA synthetase, respectively.

Activation of fatty acids is mediated by at least three forms of acyl-CoA synthetase activity: i) acetyl-CoA synthetase, which activates acetate and several other low molecular weight carboxylic acids and is found in muscle mitochondria and the cytosol of other tissues; ii) medium-chain acyl-CoA synthetase, which activates fatty acids containing between four and eleven carbon atoms (predominantly from dietary sources), and is present only in liver mitochondria; and iii) acyl CoA synthetase, which is specific for long chain fatty acids with between six and twenty carbon atoms, and is found in microsomes and the mitochondria. Proteins associated with acyl-CoA synthetase activity have been identified from many sources including bacteria, yeast, plants, mouse, and man. The activity of acyl-CoA synthetase may be modulated by phosphorylation of the enzyme by cAMP-

dependent protein kinase.

Lysophosphatidic acid (LPA) and phosphatidic acid (PA) are two phospholipids involved in signal transduction and in lipid biosynthesis in cells. LPA acyltransferase (LPAAT), also known as 1-acyl *sn*-glycerol-3-phosphate acetyltransferase (EC 2.3.1.51), catalyzes the conversion of LPA to PA, utilizing acyl-CoA as the acyl donor. This enzyme is found in microsomes and the plasma membrane and may be regulated by phosphorylation in response to interleukin-1. LPA serves as an intermediate in membrane phospholipid metabolism and is also produced by activated platelets. West, J. et al. (1997, DNA Cell Biol. 16:691-701) identified messenger RNA encoding two proteins, LPAAT- $\alpha$  and LPAAT- $\beta$ , which contain extensive sequence similarities to microbial or plant LPAAT sequences. LPAAT- $\alpha$  mRNA was detected in all tissues, with the highest expression in skeletal muscle, whereas LPAAT- $\beta$  was expressed predominantly in heart and liver tissues. Overexpression of the two cDNAs in mammalian cells led to increased LPAAT activity in cell-free extracts using an *in vitro* assay. The assay measured the conversion of fluorescently labeled LPA to PA. This increase in LPAAT activity correlated with enhancement of transcription and synthesis of tumor necrosis factor- $\alpha$  and interleukin-6 in cells upon stimulation with interleukin-1 $\beta$ , suggesting that LPAAT overexpression may amplify cellular signaling responses from cytokines (West, *supra*).

In separate experiments, recombinant protein produced in COS 7 cells exhibited LPAAT activity, with a preference for LPA as the phosphoglycerol acceptor and arachidonyl coenzyme A as the acyl donor. Northern blot analysis demonstrated that the mRNA was expressed in most human tissues including a panel of brain subregions. Expression was highest in liver and pancreas and lowest in placenta (Eberhardt, C. et al. (1997) J. Biol. Chem. 272:20299-20305). LPA may also participate in the pathophysiology of neurodegenerative processes by causing vasoconstriction as well as impairment of glutamate and glucose uptake by astrocytes. In addition, LPA is a potent promoter of tumor cell growth and invasion through activation of Ras and the Raf/mitogen-activated protein kinase pathway, stimulation of phospholipases C and D, inhibition of adenylyl cyclase, and tyrosine phosphorylation of focal adhesion proteins which accompanies actin cytoskeleton remodeling (Eberhardt, *supra*).

Another human homolog of bacterial, yeast, and plant LPAAT was found in human U937 cells. Northern blot analysis indicated high levels of expression in immune cells and in epithelium. Rapid amplification of cDNA ends revealed differentially expressed splice variants (Stamps, A.C. et al. (1997) Biochem. J. 326:455-461).

The discovery of new coenzyme A-utilizing enzymes and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of neoplastic, immunological, neurological, vesicle trafficking, and muscle disorders.

### SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, coenzyme A-utilizing enzymes, referred to collectively as "CoAEN" and individually as "CoAEN-1," "CoAEN-2," "CoAEN-3," "CoAEN-4," and "CoAEN-5." In one aspect, the invention provides a substantially purified  
5 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof. The invention also includes a polypeptide comprising an amino acid sequence that differs by one or more conservative amino acid substitutions from an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

The invention further provides a substantially purified variant having at least 90% amino acid  
10 identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1-5 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide  
15 encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof. The invention  
20 also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the  
25 polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a  
30 polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10 and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising  
35 a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10 and fragments

WO 00/26350

PCT/US99/25820

thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In another aspect, the expression vector is contained within a host cell.

5 The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing a polynucleotide of the invention under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

10 The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-5 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

15 The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of CoAEN, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

20 The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of CoAEN, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof.

## 25 BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding CoAEN.

30 Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of CoAEN.

Table 3 shows the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

35 Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding CoAEN were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze CoAEN, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"CoAEN" refers to the amino acid sequences of substantially purified CoAEN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CoAEN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CoAEN either by directly interacting with CoAEN or by acting on components of the biological pathway in which CoAEN participates.

An "allelic variant" is an alternative form of the gene encoding CoAEN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.

Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding CoAEN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CoAEN or a polypeptide with at least one functional characteristic of CoAEN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CoAEN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CoAEN. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CoAEN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CoAEN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to an amino acid sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of CoAEN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CoAEN either by directly interacting with CoAEN or by acting on components of the biological pathway in which CoAEN participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CoAEN polypeptides can be prepared using intact polypeptides or using



fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic CoAEN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.

Compositions comprising polynucleotide sequences encoding CoAEN or fragments of CoAEN may

be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

“Consensus sequence” refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the

absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of CoAEN or the polynucleotide encoding CoAEN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:6-10 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:6-10, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:6-10 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:6-10 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:6-10 and the region of SEQ ID NO:6-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-5 is encoded by a fragment of SEQ ID NO:6-10. A fragment of SEQ ID NO:1-5 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-5. For example, a fragment of SEQ ID NO:1-5 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-5. The precise length of a fragment of SEQ ID NO:1-5 and the region of SEQ ID NO:1-5 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially

WO 00/26350

PCT/US99/25820

complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2

WO 00/26350

PCT/US99/25820

Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default



SDS, and about 100 µg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

WO 00/26350

PCT/US99/25820

The term "modulate" refers to a change in the activity of CoAEN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CoAEN.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding CoAEN, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A



Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a

vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CoAEN, or fragments thereof, or CoAEN itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-

WO 00/26350

PCT/US99/25820

1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human coenzyme A-utilizing enzymes (CoAEN), the polynucleotides encoding CoAEN, and the use of these compositions for the diagnosis, treatment, or prevention of neoplastic, immunological, neurological, vesicle trafficking, and muscle disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding CoAEN. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each CoAEN were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The Incyte clones in column 5 were used to assemble the consensus nucleotide sequence of each CoAEN and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each

WO 00/26350

PCT/US99/25820

polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows the identity of each protein; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding CoAEN. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists tissue categories which express CoAEN as a fraction of total tissues expressing CoAEN. Column 3 lists diseases, disorders, or conditions associated with those tissues expressing CoAEN as a fraction of total tissues expressing CoAEN. Column 4 lists the vectors used to subclone each cDNA library.

The following fragments of the nucleotide sequences encoding CoAEN are useful in hybridization or amplification technologies to identify SEQ ID NO:6-10 and to distinguish between SEQ ID NO:6-10 and related polynucleotide sequences. The encoded polypeptide sequences are useful, for example, in generating antibodies. The useful fragments are the fragment of SEQ ID NO:6 from about nucleotide 515 to about nucleotide 607; the fragments of SEQ ID NO:7 from about nucleotide 1229 to about nucleotide 1264 and from about nucleotide 1619 to about nucleotide 1642; the fragment of SEQ ID NO:8 from about nucleotide 1258 to about nucleotide 1437; the fragment of SEQ ID NO:9 from about nucleotide 1191 to about nucleotide 1229; and the fragment of SEQ ID NO:10 from about nucleotide 77 to about nucleotide 139.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding CoAEN were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The invention also encompasses CoAEN variants. A preferred CoAEN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CoAEN amino acid sequence, and which contains at least one functional or structural characteristic of CoAEN.

The invention also encompasses polynucleotides which encode CoAEN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10, which encodes CoAEN.

The invention also encompasses a variant of a polynucleotide sequence encoding CoAEN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CoAEN. A particular aspect of the invention encompasses a variant of a

WO 00/26350

PCT/US99/25820

polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:6-10. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CoAEN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CoAEN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CoAEN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CoAEN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CoAEN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CoAEN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CoAEN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CoAEN and CoAEN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CoAEN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:6-10 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of

the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CoAEN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060).

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been

size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

5           Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate  
10   software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

          In another embodiment of the invention, polynucleotide sequences or fragments thereof  
15   which encode CoAEN may be cloned in recombinant DNA molecules that direct expression of CoAEN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CoAEN.

          The nucleotide sequences of the present invention can be engineered using methods generally  
20   known in the art in order to alter CoAEN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction  
25   sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

          In another embodiment, sequences encoding CoAEN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CoAEN itself or a fragment thereof may be synthesized using chemical methods. For  
30   example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of CoAEN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

35           The peptide may be substantially purified by preparative high performance liquid

chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins. Structures and Molecular Properties, WH Freeman, New York NY.)

5 In order to express a biologically active CoAEN, the nucleotide sequences encoding CoAEN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and  
10 polynucleotide sequences encoding CoAEN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CoAEN. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CoAEN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional  
15 transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used.  
20 (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CoAEN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning. A  
25 Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CoAEN. These include, but are not limited to, microorganisms such as bacteria transformed  
30 with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

35 In bacterial systems, a number of cloning and expression vectors may be selected depending



WO 00/26350

PCT/US99/25820

upon the use intended for polynucleotide sequences encoding CoAEN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CoAEN can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CoAEN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Hecke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CoAEN are needed, e.g. for the production of antibodies, vectors which direct high level expression of CoAEN may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CoAEN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CoAEN. Transcription of sequences encoding CoAEN may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CoAEN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CoAEN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

WO 00/26350

PCT/US99/25820

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CoAEN in cell lines is preferred. For example, sequences encoding CoAEN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CoAEN is inserted within a marker gene sequence, transformed cells containing sequences encoding CoAEN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CoAEN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates

expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CoAEN and that express CoAEN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR

5 amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CoAEN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and

10 fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CoAEN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and  
15 Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CoAEN  
20 include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CoAEN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety  
25 of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CoAEN may be cultured under  
30 conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CoAEN may be designed to contain signal sequences which direct secretion of CoAEN through a prokaryotic or eukaryotic cell membrane.

35 In addition, a host cell strain may be chosen for its ability to modulate expression of the

inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity.

- 5 Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

- 10 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CoAEN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CoAEN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CoAEN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available
- 15 affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity
- 20 purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CoAEN encoding sequence and the heterologous protein sequence, so that CoAEN may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10).
- 25 A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

- In a further embodiment of the invention, synthesis of radiolabeled CoAEN may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the
- 30 T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

- Fragments of CoAEN may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be
- 35 achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of

CoAEN may be synthesized separately and then combined to produce the full length molecule.

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CoAEN and coenzyme A-utilizing enzymes. In addition, the expression of CoAEN is closely associated with cancer, immune response, inflammation, and development; and in brain and muscle tissues. Therefore, CoAEN appears to play a role in neoplastic, immunological, neurological, vesicle trafficking, and muscle disorders. In the treatment of disorders associated with increased CoAEN expression or activity, it is desirable to decrease the expression or activity of CoAEN. In the treatment of disorders associated with decreased CoAEN expression or activity, it is desirable to increase the expression or activity of CoAEN.

Therefore, in one embodiment, CoAEN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CoAEN. Examples of such disorders include, but are not limited to, an immunological disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy,

neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a vesicle trafficking disorder, such as cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminthic, and protozoal infections; and a muscle disorder, such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, and pheochromocytoma, and myopathies including cardiomyopathy, encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, and ophthalmoplegia. A muscle disorder is defined as any impairment or alteration in the normal action of muscle and includes, but is not limited to, disorders of the blood vessels, gastrointestinal tract, heart, uterus, bladder, and skeletal muscle.

In another embodiment, a vector capable of expressing CoAEN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CoAEN including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified CoAEN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CoAEN including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CoAEN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CoAEN including, but not limited to, those listed above.

WO 00/26350

PCT/US99/25820

In a further embodiment, an antagonist of CoAEN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CoAEN. Examples of such disorders include, but are not limited to, a neoplastic disorder, such as, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and disorders such as those described above. In one aspect, an antibody which specifically binds CoAEN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CoAEN.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CoAEN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CoAEN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CoAEN may be produced using methods which are generally known in the art. In particular, purified CoAEN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CoAEN. Antibodies to CoAEN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CoAEN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to

WO 00/26350

PCT/US99/25820

CoAEN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CoAEN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CoAEN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CoAEN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for CoAEN may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CoAEN and its



WO 00/26350

PCT/US99/25820

specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CoAEN epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CoAEN. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of CoAEN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CoAEN epitopes, represents the average affinity, or avidity, of the antibodies for CoAEN. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular CoAEN epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the CoAEN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CoAEN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CoAEN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, *supra*, and Coligan et al. *supra*.)

In another embodiment of the invention, the polynucleotides encoding CoAEN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding CoAEN may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding CoAEN. Thus, complementary molecules or fragments may be used to modulate CoAEN activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CoAEN.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or

from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding CoAEN. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

5 Genes encoding CoAEN can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding CoAEN. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more  
10 with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding CoAEN. Oligonucleotides derived from the transcription  
15 initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber,  
20 B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme  
25 molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CoAEN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,  
30 GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

35 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared

by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CoAEN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of CoAEN, antibodies to CoAEN, and mimetics, agonists, antagonists, or inhibitors of CoAEN. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

WO 00/26350

PCT/US99/25820

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's

5 Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

10 Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, 15 hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar 20 solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of 25 gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

30 Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily 35 injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or



compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

15 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind CoAEN may be used for the diagnosis of disorders characterized by expression of CoAEN, or in assays to monitor patients being treated with CoAEN or agonists, antagonists, or inhibitors of CoAEN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CoAEN include methods which utilize the antibody and a label to detect CoAEN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CoAEN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CoAEN expression.

30 Normal or standard values for CoAEN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to CoAEN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CoAEN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values.

35 Deviation between standard and subject values establishes the parameters for diagnosing disease.

WO 00/26350

PCT/US99/25820

In another embodiment of the invention, the polynucleotides encoding CoAEN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CoAEN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CoAEN, and to monitor regulation of CoAEN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CoAEN or closely related molecules may be used to identify nucleic acid sequences which encode CoAEN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CoAEN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CoAEN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:6-10 or from genomic sequences including promoters, enhancers, and introns of the CoAEN gene.

Means for producing specific hybridization probes for DNAs encoding CoAEN include the cloning of polynucleotide sequences encoding CoAEN or CoAEN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CoAEN may be used for the diagnosis of disorders associated with expression of CoAEN. Examples of such disorders include, but are not limited to, a neoplastic disorder, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immunological disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema

WO 00/26350

PCT/US99/25820

nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic

5 anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other

10 extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-

15 Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders,

20 peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a vesicle trafficking disorder, such as cystic fibrosis, glucose-galactose malabsorption syndrome,

25 hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking, including acquired immunodeficiency syndrome (AIDS); allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis;

30 inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjogren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminthic, and protozoal infections; and a muscle disorder, such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy,

35 centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis,



WO 00/26350

PCT/US99/25820

dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, and pheochromocytoma, and myopathies including cardiomyopathy, encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, and ophthalmoplegia. The polynucleotide sequences encoding CoAEN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CoAEN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CoAEN may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CoAEN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CoAEN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CoAEN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CoAEN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the

development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

5 Additional diagnostic uses for oligonucleotides designed from the sequences encoding CoAEN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CoAEN, or a fragment of a polynucleotide complementary to the polynucleotide encoding CoAEN, and will be employed under optimized conditions for identification of a specific gene or  
10 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of CoAEN include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C.  
15 et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the  
20 polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

25 Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

30 In another embodiment of the invention, nucleic acid sequences encoding CoAEN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single  
35 chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price,

WO 00/26350

PCT/US99/25820

C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the  
5 Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CoAEN on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

10 In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides  
15 valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be  
20 used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CoAEN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a  
25 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CoAEN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are  
30 synthesized on a solid substrate. The test compounds are reacted with CoAEN, or fragments thereof, and washed. Bound CoAEN is then detected by methods well known in the art. Purified CoAEN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

35 In another embodiment, one may use competitive drug screening assays in which neutralizing

WO 00/26350

PCT/US99/25820

antibodies capable of binding CoAEN specifically compete with a test compound for binding CoAEN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CoAEN.

In additional embodiments, the nucleotide sequences which encode CoAEN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Attorney Docket No. PF-0622 P, filed November 3, 1998], are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A<sup>+</sup>) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## 10 II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## 25 III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel,

1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:6-10. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### IV. Northern Analysis

WO 00/26350

PCT/US99/25820

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding CoAEN occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

#### V. Extension of CoAEN Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:6-10 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one

extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with



20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:6-10 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

#### VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:6-10 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise

the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645 ) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

#### VIII. Complementary Polynucleotides

Sequences complementary to the CoAEN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CoAEN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CoAEN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CoAEN-encoding transcript.

#### IX. Expression of CoAEN

Expression and purification of CoAEN is achieved using bacterial or virus-based expression systems. For expression of CoAEN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CoAEN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CoAEN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CoAEN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther.

WO 00/26350

PCT/US99/25820

7:1937-1945.)

In most expression systems, CoAEN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CoAEN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified CoAEN obtained by these methods can be used directly in the following activity assay.

#### X. Demonstration of CoAEN Activity

CoAEN activity is determined by measuring the transfer of malonyl groups from malonyl-CoA to [<sup>3</sup>H]-labeled carnitine to yield [<sup>3</sup>H]-labeled palmitoyl carnitine according to the method of Bremer, J. (1981, Biochim. Biophys. Acta 665:628-631). Mitochondria are prepared from mammalian tissue using methods well-known in the art. The mitochondrial preparations (35-40 μg) are preincubated with various concentrations of CoAEN and malonyl-CoA (0-1 μM) in an incubation mixture containing 75 mM KCl, 50 mM mannitol, 25 mM HEPES (pH 7.3), 2 mM NaCN, 0.2 mM EGTA, 1 mM dithiothreitol, and 1% bovine serum albumin (essentially fatty acid free) at 30°C for 3 minutes. Following this incubation, 0.1 μCi of L-[methyl-<sup>3</sup>H]carnitine is added to give a final L-carnitine concentration of 200 μM, and the incubation is continued for a further 6 minutes. The reaction is stopped by adding 100 μl of 10 N HCl. [<sup>3</sup>H]palmitoyl carnitine is extracted with butanol and counted using standard liquid scintillation procedures. The amount of [<sup>3</sup>H]palmitoyl carnitine recovered is proportional to the amount of CoAEN in the sample.

Alternatively, CoAEN activity is determined according to the method of Aguado and Campbell (1998, J. Biol. Chem. 273:4096-4105). CoAEN activity is determined by measuring the reaction of the thiol group of the released CoA with 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) spectrophotometrically which results in an increase in absorbance at 413 nm. A suitable mammalian cell line, such as COS7, is transiently transfected with a vector containing CoAEN. After a suitable period of incubation with cell culture media to allow transient expression of CoAEN, the cells are collected and homogenized using methods well-known in the art. Cell homogenates are then incubated with buffered substrates. A typical incubation mixture consists of 100 mM Tris-HCl, pH 7.4, 1 mM DTNB, 50 μM LPA (oleoyl-*sn*-glycerol 3-phosphate), 10-45 μM acyl-CoA, and 50-150 μg

WO 00/26350

PCT/US99/25820

of cell homogenate in a total volume of 1 ml. DTNB is added as a 0.01 M solution in 0.1 M potassium phosphate buffer, pH 7.0. The reaction is initiated by the addition of acyl-CoA after preincubation of the sample with all of the other components for 2 minutes. A molar absorbance of 13,600 units is used to calculate the coupling of released CoA with DTNB. The increase in absorbance at 413 nm is proportional to the amount of CoAEN in the sample.

#### XI. Functional Assays

CoAEN function is assessed by expressing the sequences encoding CoAEN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CoAEN on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CoAEN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CoAEN and other genes of interest can be analyzed by northern analysis or microarray techniques.

**XII. Production of CoAEN Specific Antibodies**

CoAEN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CoAEN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CoAEN activity by, for example, binding the peptide or CoAEN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

**XIII. Purification of Naturally Occurring CoAEN Using Specific Antibodies**

Naturally occurring or recombinant CoAEN is substantially purified by immunoaffinity chromatography using antibodies specific for CoAEN. An immunoaffinity column is constructed by covalently coupling anti-CoAEN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CoAEN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CoAEN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CoAEN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CoAEN is collected.

**XIV. Identification of Molecules Which Interact with CoAEN**

CoAEN, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CoAEN, washed, and any wells with labeled CoAEN complex are assayed. Data obtained using different concentrations of CoAEN are used to calculate values for the number, affinity, and association of CoAEN with the candidate molecules.

**WO 00/26350**

**PCT/US99/25820**

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

- 5 Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	CDNA Fragments
1	6	1580751	DUODNOT01	391491H1 (TMLR2DT01), 1580751H1 (DUODNOT01), 3137868H1 (SMCCNOT01), 4727709H1 (GBLADIT01)
2	7	1627889	COLNPOT01	938735X12 (CERVNOT01), 1627889H1 (COLNPOT01), 1849450F6 (LUNGFET03), 2963717T6 (SCORNOT04), 3969842F6 (PROSTUT10), 4192357H1 (BRAPDIT01)
3	8	1965888	BRSTNOT04	2744849H1 (LUNGTUT11), 1452166F1 (PENITUT01), 629493X12 (KIDNNOT05), 1441289R1 (THYRNOT03), 1965888H1 (BRSTNOT04), 1965888T6 (BRSTNOT04), SBEA07024F3
4	9	2816341	BRSTNOT14	155601R1 (THP1PLB02), 1432393T1 (BEPINON01), 1456229R1 (COLNFET02), 1599161X17C1 (BLADNOT03), 1610553X11 (COLNTUT06), 1610553X15 (COLNTUT06), 1700479F6 (BLADTUT05), 2170450F6 (ENDCNOT03), 2816341H1 (BRSTNOT14), 3751460H1 (UTRSNOT18), 3803729H1 (BLADTUT03)
5	10	3324214	PTHYNOT03	141024R1 (TLYMNOR01), 2586085F6 (BRAITUT22), 3324214F7 (PTHYNOT03), 3324214H1 (PTHYNOT03), SBUA05484D1.comp

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
1	100	T43 Y40	N51	D37-G44 D87-P98 S53-G56	Acetyl-CoA synthetase	Blast, BLOCKS, PRINTS, Motifs
2	159	S95 T96 S131 T155 S90 S91 S100		S2-G5 S61-G64 G3-L13	Carnitine palmitoyltransferase I	Blast, BLOCKS, Motifs
3	215	S120 S57 T173 S5 T86 T97 T115 T209		K41-E52 L100-A135	AU-binding protein/enoyl-CoA hydratase	Blast, BLOCKS
4	720	S368 S47 S62 S136 T168 T212 T243 T356 T485 T546 T644 S6 T11 S52 S361 T459 T494 T519 Y516	N203	D280-V297 I18-S45	Acyl-CoA synthetase	Blast, BLOCKS, PRINTS, Motifs, HMM, Pfam, PROFILESCAN
5	456	S136 T306 S338 T86 S87 T125 T230 S250 T433 T58 S100 T132 S157 S363	N247 N327 N328 N362	N152-L189 M1-G37 S9-V32	Lysophosphatidic acid acyltransferase	Blast, BLOCKS, PRINTS, HMM, SPScan



Table 3

SEQ ID NO:	Tissue Expression (Fraction of Total)	Disease, Disorder, or Condition (Fraction of Total)	Vector
6	Gastrointestinal (0.600) Cardiovascular (0.200) Reproductive (0.200)	Cancer (0.600) Cell Proliferative (0.200) Trauma (0.200)	pINCY
7	Nervous (0.444) Developmental (0.222) Gastrointestinal (0.111) Reproductive (0.111) Skeletal Muscle (0.111)	Cancer (0.333) Cell Proliferative (0.222) Trauma (0.222)	pINCY
8	Reproductive (0.358) Gastrointestinal (0.151) Nervous (0.142)	Cancer (0.585) Inflammation (0.151) Fetal (0.104)	PSPORT1
9	Reproductive (0.333) Nervous (0.150) Gastrointestinal (0.100)	Cancer (0.492) Inflammation (0.250) Cell Proliferative (0.192)	pINCY
10	Reproductive (0.306) Gastrointestinal (0.143) Hematopoietic/Immune (0.143)	Cancer (0.490) Inflammation (0.263) Cell Proliferative (0.143)	pINCY

Table 4

SEQ ID NO:	Clone ID	Library	Library Comment
6	1580751	DUODNOT01	Library was constructed using RNA isolated from duodenal tissue obtained from a 41-year-old Caucasian female during a radical pancreaticoduodenectomy. Family history included benign hypertension and malignant skin neoplasm.
7	1627889	COLNPOT01	Library was constructed using RNA isolated from colon polyp tissue removed from a 40-year-old Caucasian female during a total colectomy. Pathology indicated an inflammatory pseudopolyp; this tissue was associated with a focally invasive grade 2 adenocarcinoma and multiple tubovillous adenomas. Patient history included a benign neoplasm of the bowel.
8	1965888	BRSTNOT04	Library was constructed using RNA isolated from breast tissue removed from a 62-year-old East Indian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 ductal carcinoma. Patient history included benign hypertension, hyperlipidemia, and hematuria. Family history included cerebrovascular and cardiovascular disease, hyperlipidemia, and liver cancer.
9	2816341	BRSTNOT14	Library was constructed using RNA isolated from breast tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. Patient history included a benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.
10	3324214	PTHYNOT03	Library was constructed using RNA isolated from the left parathyroid tissue of a 69-year-old Caucasian female during a partial parathyroidectomy. Pathology indicated hyperplasia. The patient presented with primary hyperparathyroidism.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLOCKS IMPROVED Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger, and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater, Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

WO 00/26350

PCT/US99/25820

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 and fragments thereof.

2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.

3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.

4. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 3.

5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.

6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.

7. A method for detecting a polynucleotide, the method comprising the steps of:

(a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and

(b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.

8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.

9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10 and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 70% polynucleotide sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary

**WO 00/26350**

**PCT/US99/25820**

to the polynucleotide of claim 9.

12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

5 13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:

a) culturing the host cell of claim 13 under conditions suitable for the expression

of the polypeptide; and

10 b) recovering the polypeptide from the host cell culture.

15 15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

18. A purified antagonist of the polypeptide of claim 1.

20 19. A method for treating or preventing a disorder associated with decreased expression or activity of CoAEN, the method comprising administering to a subject in need of such treatment an effective amount of the pharmaceutical composition of claim 15.

25 20. A method for treating or preventing a disorder associated with increased expression or activity of CoAEN, the method comprising administering to a subject in need of such treatment an effective amount of the antagonist of claim 18.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
11 May 2000 (11.05.2000)

PCT

(10) International Publication Number  
WO 00/26350 A3

(51) International Patent Classification<sup>7</sup>: C12N 9/00,  
15/52, C07K 16/40, A61K 38/43

[US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). **KASER, Matthew, R.** [GB/US]; 4793 Ewing Road, Castro Valley, CA 94546 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).

(21) International Application Number: PCT/US99/25820

(22) International Filing Date:  
3 November 1999 (03.11.1999)

(74) Agents: **BILLINGS, Lucy, J.** et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/185,217 3 November 1998 (03.11.1998) US  
Not furnished 3 November 1998 (03.11.1998) US

(81) Designated States (*national*): AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW.

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 09/185,217 (CIP)  
Filed on 3 November 1998 (03.11.1998)  
US Not furnished (CIP)  
Filed on 3 November 1998 (03.11.1998)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **INCYTE PHARMACEUTICALS, INC.** [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

**Published:**

- With international search report.
- Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **CORLEY, Neil, C.** [US/US]; 1240 Dale Avenue #30, Mountain View, CA 94040 (US). **GUEGLER, Karl, J.** [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). **GORGONE, Gina, A.** [US/US]; 1253 Pinecrest Drive, Boulder Creek, CA 95006 (US). **AZIMZAI, Valda**

(88) Date of publication of the international search report:  
7 December 2000

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: COENZYME A-UTILIZING ENZYMES

(57) Abstract: The invention provides human coenzyme A-utilizing enzymes (CoAEN) and polynucleotides which identify and encode CoAEN. The invention also provides expression vectors, host cells, antibodies, agonists and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CoAEN.

WO 00/26350 A3

Docket No.: PF-0622 USN

**DECLARATION AND POWER OF ATTORNEY FOR  
UNITED STATES PATENT APPLICATION**

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,  
and

I believe that I am the original, first and sole inventor (if only one name is listed below)  
or an original, first and joint inventor (if more than one name is listed below) of the subject  
matter which is claimed and for which a United States patent is sought on the invention entitled

**COENZYME A-UTILIZING ENZYMES**

the specification of which:

  /   / is attached hereto.

  /   / was filed on \_\_\_\_\_ as application Serial No. \_\_\_\_\_ and if this box  
contains an X   /   /, was amended on \_\_\_\_\_.

  / X /   / was filed as Patent Cooperation Treaty international application No. PCT/US99/25820  
on 03 November, 1999, if this box contains an X   /   /, was amended on under Patent Cooperation  
Treaty Article 19 on \_\_\_\_\_ 2001, and if this box contains an X   /   /, was amended on \_\_\_\_\_.

I hereby state that I have reviewed and understand the contents of the above-identified  
specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of  
this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any  
foreign application(s) for patent or inventor's certificate indicated below and of any Patent  
Cooperation Treaty international applications(s) designating at least one country other than the  
United States indicated below and have also identified below any foreign application(s) for  
patent or inventor's certificate and Patent Cooperation Treaty international application(s)  
designating at least one country other than the United States for the same subject matter and  
having a filing date before that of the application for said subject matter the priority of which is  
claimed:



Docket No.: PF-0622 USN

Country	Number	Filing Date	Priority Claimed
_____	_____	_____	// Yes // No
_____	_____	_____	// Yes // No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application		Status (Pending,
Serial No.	Filed	Abandoned, Patented)
60/240,921	03 November 1998	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application		Status (Pending,
Serial No.	Filed	Abandoned, Patented)
_____	_____	_____

I hereby appoint the following:

(14) Lucy J. Billings	Reg. No. <u>36,749</u>
Michael C. Cerrone	Reg. No. <u>39,132</u>
Diana Hamlet-Cox	Reg. No. <u>33,302</u>
Richard C. Ekstrom	Reg. No. <u>37,027</u>
Barrie D. Greene	Reg. No. <u>46,740</u>
Matthew R. Kaser	Reg. No. <u>44,817</u>
Lynn E. Murry	Reg. No. <u>42,918</u>
Shirley A. Recipon	Reg. No. <u>47,016</u>
Susan K. Sather	Reg. No. <u>44,316</u>
Michelle M. Stempien	Reg. No. <u>41,327</u>
David G. Streeter	Reg. No. <u>43,168</u>
Stephen Todd	Reg. No. <u>47,139</u>
Christopher Turner	Reg. No. <u>45,167</u>
P. Ben Wang	Reg. No. <u>41,420</u>

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0622 USN

**LEGAL DEPARTMENT**  
**INCYTE GENOMICS, INC.**  
**3160 PORTER DRIVE, PALO ALTO, CA 94304**

TEL: 650-855-0555

FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

1-00  
Sole Inventor or  
First Joint Inventor:

Full name:

Y. TOM TANG

Signature:

U. Tom Tang

Date:

Feb. 27, 2001

Citizenship

People's Republic of China USA

Residence:

San Jose, California 95118 CA

P.O. Address:

4230 Ranwick Court  
San Jose, California 95118

Docket No.: PF-0622 USN

2-00  
Second Joint Inventor:

Full name:

NEIL C. CORLEY

Signature:

*Neil C. Corley*

Date:

MARCH 5, 2001

Citizenship

United States

Residence:

Castro Valley, California 94552 CA

P.O. Address:

20426 Crow Creek Road  
Castro Valley, California 94552

3-00  
Third Joint Inventor:

Full name:

KARL J. GUEGLER

Signature:

*K. J. Gugler*

Date:

02/02, 2001

Citizenship

Switzerland

Residence:

Menlo Park, California 94025 CA

P.O. Address:

1048 Oakland Avenue  
Menlo Park, California 94025

4-00  
Fourth Joint Inventor:

Full name:

GINA A. GORGONE

Signature:

*Gina A. Gorgone*

Date:

2/5, 2001

Citizenship

United States

Residence:

Boulder Creek, California 95006 CA

P.O. Address:

1253 Pinecrest Drive  
Boulder Creek, California 95006

Docket No.: PF-0622 USN

5-00

Fifth Joint Inventor:

Full name:

YALDA AZIMZAI

Signature:

*Yalda Azimzai*

Date:

2nd February, 2001

Citizenship

United States of America

Residence:

Castro Valley, California 94552 CA

P.O. Address:

5518 Boulder Canyon Drive  
Castro Valley, California 94552

6-00

Sixth Joint Inventor:

Full name:

MATTHEW R. KASER

Signature:

*Matthew R. Kaser*

Date:

2nd February, 2001

Citizenship

United Kingdom

Residence:

Castro Valley, California 94546 CA

P.O. Address:

4793 Ewing Road  
Castro Valley, California 94546

7-00

Seventh Joint Inventor:

Full name:

HENRY YUE

Signature:

*Henry Yue*

Date:

March 02, 2001

Citizenship

United States

Residence:

Sunnyvale, California 94087 CA

P.O. Address:

826 Lois Avenue  
Sunnyvale, California 94087

## SEQUENCE LISTING

&lt;110&gt; INCYTE PHARMACEUTICALS, INC.

TANG, Y. Tom

CORLEY, Neil C.

GUEGLER, Karl J.

GORGONE, Gina A.

AZIMZAI, Yalda

KASER, Matthew R.

YUE, Henry

&lt;120&gt; COENZYME A-UTILIZING ENZYMES

&lt;130&gt; PF-0622 PCT

&lt;140&gt; To Be Assigned

&lt;141&gt; Herewith

&lt;150&gt; 09/185,217; unassigned

&lt;151&gt; 1998-11-03; 1998-11-03

&lt;160&gt; 10

&lt;170&gt; PERL Program

&lt;210&gt; 1

&lt;211&gt; 100

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1580751CD1

&lt;400&gt; 1

Met	Lys	Gly	Ala	Thr	Thr	Asn	Ile	Cys	Tyr	Asn	Val	Leu	Asp	Arg
1				5					10					15
Asn	Val	His	Glu	Lys	Lys	Leu	Gly	Asp	Lys	Val	Ala	Phe	Tyr	Trp
			20						25					30
Pro	Cys	Gln	Arg	Asp	Gln	Asp	Gly	Tyr	Tyr	Trp	Ile	Thr	Gly	Arg
			35						40					45
Ile	Asp	Asp	Met	Leu	Asn	Val	Ser	Gly	Glu	Gly	Gln	Gly	Pro	Pro
			50						55					60
Ser	His	Leu	Ile	Asn	Ser	Ala	Pro	Leu	Thr	Thr	Pro	Ser	Arg	Ser
			65						70					75
Leu	Pro	Gln	Glu	Pro	Arg	Ser	Val	Leu	Trp	Pro	Asp	His	Val	Leu
			80						85					90
Ser	Val	Ala	Phe	Ser	Ser	Gly	Pro	Arg	Phe					
			95						100					

&lt;210&gt; 2

&lt;211&gt; 159

WO 00/26350

PCT/US99/25820

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1627889CD1

<400> 2

Met	Ser	Gly	Gln	Gly	Val	Asp	Arg	His	Leu	Phe	Ala	Leu	Tyr	Ile
1				5					10					15
Val	Ser	Arg	Phe	Leu	His	Leu	Gln	Ser	Pro	Phe	Leu	Thr	Gln	Val
				20					25					30
His	Ser	Glu	Gln	Trp	Gln	Leu	Ser	Thr	Ser	Gln	Ile	Pro	Val	Gln
				35					40					45
Gln	Met	His	Leu	Phe	Asp	Val	His	Asn	Tyr	Pro	Asp	Tyr	Val	Ser
				50					55					60
Ser	Gly	Gly	Gly	Phe	Gly	Pro	Ala	Asp	Asp	His	Gly	Tyr	Gly	Val
				65					70					75
Ser	Tyr	Ile	Phe	Met	Gly	Asp	Gly	Met	Ile	Thr	Phe	His	Ile	Ser
				80					85					90
Ser	Lys	Lys	Ser	Ser	Thr	Lys	Thr	Asp	Ser	His	Arg	Leu	Gly	Gln
				95					100					105
His	Ile	Glu	Asp	Ala	Leu	Leu	Asp	Val	Ala	Ser	Leu	Phe	Gln	Ala
				110					115					120
Gly	Gln	His	Phe	Lys	Arg	Arg	Phe	Arg	Gly	Ser	Gly	Lys	Glu	Asn
				125					130					135
Ser	Arg	His	Arg	Cys	Gly	Phe	Leu	Ser	Arg	Gln	Thr	Gly	Ala	Ser
				140					145					150
Lys	Ala	Ser	Met	Thr	Ser	Thr	Asp	Phe						
				155										

<210> 3

<211> 215

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1965888CD1

<400> 3

Met	Asn	Arg	Pro	Ser	Ala	Arg	Asn	Ala	Leu	Gly	Asn	Val	Phe	Val
1				5					10					15
Ser	Glu	Leu	Leu	Glu	Thr	Leu	Ala	Gln	Leu	Arg	Glu	Asp	Arg	Gln
				20					25					30
Val	Arg	Val	Leu	Leu	Phe	Arg	Ser	Gly	Val	Lys	Gly	Val	Phe	Cys
				35					40					45
Ala	Gly	Ala	Asp	Leu	Lys	Glu	Arg	Glu	Gln	Met	Ser	Glu	Ala	Glu
				50					55					60
Val	Gly	Val	Phe	Val	Gln	Arg	Leu	Arg	Gly	Leu	Met	Asn	Asp	Ile
				65					70					75
Ala	Ser	Ser	Ala	Val	Met	Gly	Leu	Ile	Glu	Thr	Thr	Arg	Gly	Leu
				80					85					90
Leu	Pro	Gly	Ala	Gly	Gly	Thr	Gln	Arg	Leu	Pro	Arg	Cys	Leu	Gly







WO 00/26350

PCT/US99/25820

	605		610		615
Ser Tyr His Ser Tyr Val Ile Gly Phe Val Val Pro Asn Gln Lys					
	620		625		630
Glu Leu Thr Glu Leu Ala Arg Lys Lys Gly Leu Lys Gly Thr Trp					
	635		640		645
Glu Glu Leu Cys Asn Ser Cys Glu Met Glu Asn Glu Leu Leu Lys					
	650		655		660
Val Leu Ser Glu Ala Ala Ile Ser Ala Ser Leu Glu Lys Phe Glu					
	665		670		675
Ile Leu Val Lys Ile Arg Leu Ser Pro Glu Pro Trp Thr Pro Glu					
	680		685		690
Thr Gly Leu Val Thr Asp Ala Phe Lys Leu Lys Arg Lys Glu Leu					
	695		700		705
Lys Thr His Tyr Gln Ala Asp Ile Glu Arg Met Tyr Gly Arg Lys					
	710		715		720

<210> 5

<211> 456

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3324214CD1

<400> 5

Met Phe Leu Leu Leu Thr Phe Asp Ser Leu Ile Val Asn Leu Leu		
1 5 10 15		
Gly Ile Ser Leu Thr Val Leu Phe Thr Leu Leu Leu Val Phe Ile		
20 25 30		
Ile Val Pro Ala Ile Phe Gly Val Ser Phe Gly Ile Arg Lys Leu		
35 40 45		
Tyr Met Lys Ser Leu Leu Lys Ile Phe Ala Trp Ala Thr Leu Arg		
50 55 60		
Met Glu Arg Gly Ala Lys Glu Lys Asn His Gln Leu Tyr Lys Pro		
65 70 75		
Tyr Thr Asn Gly Ile Ile Ala Lys Asp Pro Thr Ser Leu Glu Glu		
80 85 90		
Glu Ile Lys Glu Ile Arg Arg Ser Gly Ser Ser Lys Ala Leu Asp		
95 100 105		
Asn Thr Pro Glu Phe Glu Leu Ser Asp Ile Phe Tyr Phe Cys Arg		
110 115 120		
Lys Gly Met Glu Thr Ile Met Asp Asp Glu Val Thr Lys Arg Phe		
125 130 135		
Ser Ala Glu Glu Leu Glu Ser Trp Asn Leu Leu Ser Arg Thr Asn		
140 145 150		
Tyr Asn Phe Gln Tyr Ile Ser Leu Arg Leu Thr Val Leu Trp Gly		
155 160 165		
Leu Gly Val Leu Ile Arg Tyr Cys Phe Leu Leu Pro Leu Arg Ile		
170 175 180		
Ala Leu Ala Phe Thr Gly Ile Ser Leu Leu Val Val Gly Thr Thr		
185 190 195		
Val Val Gly Tyr Leu Pro Asn Gly Arg Phe Lys Glu Phe Met Ser		
200 205 210		

WO 00/26350

PCT/US99/25820

Lys	His	Val	His	Leu	Met	Cys	Tyr	Arg	Ile	Cys	Val	Arg	Ala	Leu
				215					220					225
Thr	Ala	Ile	Ile	Thr	Tyr	His	Asp	Arg	Glu	Asn	Arg	Pro	Arg	Asn
				230					235					240
Gly	Gly	Ile	Cys	Val	Ala	Asn	His	Thr	Ser	Pro	Ile	Asp	Val	Ile
				245					250					255
Ile	Leu	Ala	Ser	Asp	Gly	Tyr	Tyr	Ala	Met	Val	Gly	Gln	Val	His
				260					265					270
Gly	Gly	Leu	Met	Gly	Val	Ile	Gln	Arg	Ala	Met	Val	Lys	Ala	Cys
				275					280					285
Pro	His	Val	Trp	Phe	Glu	Arg	Ser	Glu	Val	Lys	Asp	Arg	His	Leu
				290					295					300
Val	Ala	Lys	Arg	Leu	Thr	Glu	His	Val	Gln	Asp	Lys	Ser	Lys	Leu
				305					310					315
Pro	Ile	Leu	Ile	Phe	Pro	Glu	Gly	Thr	Cys	Ile	Asn	Asn	Thr	Ser
				320					325					330
Val	Met	Met	Phe	Lys	Lys	Gly	Ser	Phe	Glu	Ile	Gly	Ala	Thr	Val
				335					340					345
Tyr	Pro	Val	Ala	Ile	Lys	Tyr	Asp	Pro	Gln	Phe	Gly	Asp	Ala	Phe
				350					355					360
Trp	Asn	Ser	Ser	Lys	Tyr	Gly	Met	Val	Thr	Tyr	Leu	Leu	Arg	Met
				365					370					375
Met	Thr	Ser	Trp	Ala	Ile	Val	Cys	Ser	Val	Trp	Tyr	Leu	Pro	Pro
				380					385					390
Met	Thr	Arg	Glu	Ala	Asp	Glu	Asp	Ala	Val	Gln	Phe	Ala	Asn	Arg
				395					400					405
Val	Lys	Ser	Ala	Ile	Ala	Arg	Gln	Gly	Gly	Leu	Val	Asp	Leu	Leu
				410					415					420
Trp	Asp	Gly	Gly	Leu	Lys	Arg	Glu	Lys	Val	Lys	Asp	Thr	Phe	Lys
				425					430					435
Glu	Glu	Gln	Gln	Lys	Leu	Tyr	Ser	Lys	Met	Ile	Val	Gly	Asn	His
				440					445					450
Lys	Asp	Arg	Ser	Arg	Ser									
				455										

<210> 6

<211> 687

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> 63

<223> a or g or c or t, unknown, or other

<220>

<221> misc\_feature

<223> Incyte ID No: 1580751CB1

<400> 6

```

cggttctagg aacttgacgt gatggggctt cctgaggagc ggggccggag cggcagcggg 60
ganccggggc caggaggaag ctggagccgg aggcggggcg cggagttggt ctccgccgcc 120
cgaggtcagc cgctccgcgc acgtccctc gctgcagcgc taccgcgagc tgcaccggcg 180
ctccgtggag gagccgcggg aattctgggg agacattgcc aaggaatttt actggaagac 240

```

WO 00/26350

PCT/US99/25820

```
tccatgccct ggcccattcc ttcggtacaa ctttgatgtg actaaagggg aaatcttcat 300
tgagtggatg aaaggagcaa ctaccaacat ctgctacaat gtactggatc gaaatgtcca 360
tgagaaaaag cttggagata aagttgcttt ttactggcct tgccagcggg accaggatgg 420
ctattactgg atcaactggca ggattgatga catgctcaat gtatctggtg agggccaggg 480
gccaccttcc catcttatta actctgctcc tctgacaaca cccagccgaa gccttccgca 540
agagcccagg agtgtccttt ggccagacca tgtactaagt gtagcattca gttctggggc 600
cagggttttag aggagtaatg aacactggaa tgtgcctaga gtaggggaac ttggatggtg 660
gggtgactga aagcatgtca tctgagg                                     687
```

<210> 7

<211> 1803

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1627889CB1

<400> 7

```
caaattttca tttttaatag agatgggttc accaagggtg caaggctggt tttgattccc 60
gatcccagaa ccacctgccc tgggcccccc aaagtgcggg atctacaggc atgagcccc 120
gtgctcagac taaagctttt ctttatacgg gggttcatgt tggtaaaaat ccattgaatt 180
atatacgtag gatttttgaa cttatttcta taaagtttta tatttcaata aaaagcttaa 240
agatatatat atattatttt ccatacatga caagtattgt atcatatata ctatttttga 300
acttattcct ataaaatggt atatttcaat aaaaactgac agatatatta cattattttc 360
catccatgac aagtattatt atatcataca tgctattttt tttttttttt ttttttttga 420
gatggagttt cgcttggttc ccaggctgga gtgcaatggc gccatctcgg cccaccgcaa 480
cctctgcctc ccagggttcaa gcgattctcc tgccctcagcc tcttgagtag ctgggactac 540
agcatgcgcc accataccca gctaattttg tgtttttagt agagatgggg tttctccata 600
ctggtcaggg tgggtctcaa ctcctgacct cagggtgatcc gcctgcctca gcctcccaaa 660
gtgctgggat tacaggcgtg accaatgcac ctggttgga ctcattttt acaacgtaac 720
tctgcccatt taacctcttt gtgacttggt accttccctt gcacccctgt accctctctg 780
cccagatecc acctgccttg cactcttct ccttgccct acacctagcc ccgcacctac 840
aaggatttgt ggttccatgg aaggagtttg gactctgggg ccagacacac ctaggaaacc 900
gcctggctct cctggctgt gtgaacctgg ccaaatgatt tccctctct aagcctcagt 960
tccccatctg taaaatgggg ttgatattcc cacttgagc ggatgtggca aactcagttg 1020
aggccagatg tgccgcagcc agccctaagt cgacttcttg tctttccatg acctgtgacc 1080
tccctgggga ctgcaggacc cacagtgcct cgccctgttc cgcgtggcag tggacaagca 1140
ccaggctctg ctgaaggcag ccattgagcg gcaggaggtt gaccgccacc tgtttgcgct 1200
gtacatctg tcccgattcc tccacctgca gtgcaccttc ctgacccagg tccattcgga 1260
gcagtggcag ctgtccacca gccagatccc tgttcagcaa atgcatctgt ttgacgtcca 1320
caattaccgg gactatgttt cctcaggcgg tggattcggg cctgctgatg accatggtta 1380
tgggtgtttct tatactttca tgggggatgg catgatcacc ttccacatct ccagcaaaaa 1440
atcaagcaca aaaacggatt cccacaggct ggggcagcac attgaggacg cactgctgga 1500
tgtggcctcc ctgttccagg cgggacagca ttttaagcgc cggttcagag ggtcagggaa 1560
ggagaactcc aggcacaggt gtggatttct ctcccgcag actggggcct ccaaggcctc 1620
aatgacatcc accgacttct gactccttcc agcaggcagc tggcctctcc aaggaataag 1680
ggtgaaattg ccacagctgg ctgacacagg acaggggcaa ctggtttggc aacccccacat 1740
ccaggccaat aaagatgtgt gagctggaca cttataaaaa aaacaaatat aaaaaaaaaa 1800
aaa                                             1803
```

<210> 8

<211> 1340

WO 00/26350

PCT/US99/25820

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1965888CB1

<400> 8

```
ccagcctgcg ccccgatgct ggcggttctg tgcctcctgc gcccttgag gcccttcgg 60
gcccgcggct gcgtttccga cggggcgggc gggggctcag agatccaagt gcgcgccctg 120
gcggggtccg accaagttta caagtccttc cactttctct ctgaggcaga aagagcaagg 180
gtttttctct ccattttatg gttgggaaaa ttgaggcctg cctgagtgtg tgacttgtgg 240
caagtcactc tggatcatca gggcagaggc tcccagatc ccaggcctcc tgcctccagt 300
ccccagcccg cagcccagga ttaggcagag ccagctgctt tcccgtaggt gccctgactc 360
cttacaggga tcatgagat tctgatgaac agaccttctg cccgcaatgc cttgggggaat 420
gtcttcgtca gtgagctgct ggaaactctg gccagctgc gggaggaccg gcaagtgcgt 480
gtcctgctct tcagaagtgg agtgaagggc gtgttctgtg cagggtgcaga cctgaaggag 540
cggaacaga tgagtgaagc agaggtgggg gtgtttgtcc agcgactccg gggcctgatg 600
aatgacatcg ctctctcggc agtcatggga ctgattgaga ccacgcgagg gctcctcccg 660
ggggcaggag ggactcagag gctgccccgt tgtctggggg tggcctggc gaaggagctc 720
atcttcacgg gccgacgact gagtggaaact gaggccacg tactggggct ggtgaatcac 780
gctgtggccc agaacgagga gggggacgcc gcctaccagc gggcacgagc actggccag 840
gagatcctgc cccaggcccc cattgccgtg cggctgggca aagtagccat tgaccgagga 900
acggaggtgg acattgcac tgggatggcc attgaaggga tgtgctatgc ccagaatatt 960
ccaaccggg accggctaga gggcatggca gccttcaggg agaagcggac tcccaaattt 1020
gttggcaaat gaccccat ttaacctca gcatgggaga tgcattgcct gaagagcagg 1080
atccagaagg aagatttgtg gccagattgc ctcatcatt tcacctctcc agacttccat 1140
ttcttcacaa ggatgatgat ggaaataaaa tgactggcgt gatgcctgga accaagggtg 1200
tgatcctacc acctactgct accttcctta gcttcaccct ggctagaaat aatcacgagg 1260
gttgggtttg ctttggaaaa tgctgtctc tctacttgaa tgataaagaa ttaaattaga 1320
tctctctgaa aaaaaaaaaa 1340
```

<210> 9

<211> 4027

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2816341CB1

<400> 9

```
gcgcggggt gcgacactgc agttgtctac gcggccgggg cggggacgag gaggcgttgg 60
acgggggtcgc atacgttcgt cccctcgcat tgccgccccg acagctgcgc caggatcccc 120
gggcggcggc gcggggcggt aacgctctgg ggctcagcca ggctgcgcg ggcccgatgc 180
cggaggaacc cggactccgg cgtagcggtt ttgacacaag ggcgcatac ttcaaagcac 240
ctagtacctc ctaccattgt caactgatac agaattcggt gttgggaagg actggggaaa 300
cagctgtaac atttgccacc ctcaagaagt gctggtcctg tgtcacacca ccttagcctc 360
ttgatcgagg aagattctcg ctgaagtctg ttaattctac tttttgagta cttatgaata 420
accacgtgtc ttcaaaacca tctaccatga agctaaaaca taccatcaac cctattcttt 480
tatattttat acattttcta atatcacttt atactatttt aacatacatt ccgtttttatt 540
ttttctccga gtcaagacaa gaaaaatcaa accgaattaa agcaaagcct gtaaattcaa 600
aacctgattc tgcatacaga tctgttaata gtttggatgg tttggcttca gtattatacc 660
```

WO 00/26350

PCT/US99/25820

ctggatgtga	tacttttagat	aaagttttta	catatgcaaa	aaacaaat	aagaacaaaa	720
gactcttggg	aacacgtgaa	gttttaaatg	aggaagatga	agtacaacca	aatggaaaaa	780
tttttaaaaa	ggttattctt	ggacagtata	attggctttc	ctatgaagat	gtctttgttc	840
gagcctttta	ttttggaaat	ggattacaga	tgttgggtca	gaaaccaaag	accaacatcg	900
ccatcttctg	tgagaccagg	gccgagtggg	tgatagctgc	acaggcgtgt	tttatgtata	960
atcttcagct	tgttacatta	tatgccactc	taggaggtcc	agccattgtt	catgcattaa	1020
atgaaacaga	ggtgaccaac	atcattacta	gtaaagaact	cttacaacaa	aagttgaagg	1080
atatagtttc	tttgggtccc	cgcctgcggc	acatcatcac	tgttgatgga	aagccaccga	1140
cctgggtccga	gttccccaa	ggcatcattg	tgcataccat	ggctgcagtg	gaggccctgg	1200
gagccaaggg	cagcatggaa	aaccaacctc	atagcaaacc	attgccctca	gatattgcag	1260
taatcatgta	cacaagtggg	tccacaggac	ttccaaaggg	agtcatgatc	tcacatagta	1320
acattattgc	tggtataact	gggatggcag	aaaggattcc	agaactagga	gaggaagatg	1380
tctacattgg	atatttgcct	ctggcccatg	ttctagaatt	aagtgtctgag	cttgtctgtc	1440
tttctcagg	atgcgcgcat	ggttactctt	caccacagac	tttagcagat	cagtcttcaa	1500
aaattaaaaa	aggaagcaaa	ggggatacat	ccatgttgaa	accaacactg	atggcagcag	1560
ttccggaaat	catggatcgg	atctacaaaa	atgtcatgaa	taaagtcagt	gaaatgagta	1620
gttttcaacg	taatctgttt	attctggcct	ataattacaa	aatggaacag	atttcaaaag	1680
gacgtaatac	tccactgtgc	gacagctttg	ttttccggaa	agttcgaagc	ttgctagggg	1740
gaaatatctg	tctcctgttg	tgtgggtggg	ctccactttc	tgcaaccacg	cagcgattca	1800
tgaacatctg	tttctgtctg	cctgttgggt	agggatacgg	gctcactgaa	tctgtctggg	1860
ctggaacaat	ttccgaagtg	tgggactaca	atactggcag	agtgggagca	ccattagttt	1920
gctgtgaaat	caaattaaaa	aactgggagg	aaggtggata	ctttaatact	gataagccac	1980
accccgaggg	tgaaattctt	attggggggc	aaagtgtgac	aatgggggtac	tacaaaaatg	2040
aagcaaaaac	aaaagctgat	ttctttgaag	atgaaaatgg	acaaaggtgg	ctctgtactg	2100
gggatattgg	agagtgtgaa	cccgatggat	gcttaaagat	tattgatcgt	aaaaaggacc	2160
ttgtaaaaac	acaggcaggg	gaatatgttt	ctcttgggaa	agtagaggca	gctttgaaga	2220
atcttccact	agtagataac	atctgtgcat	atgcaaacag	ttatcattct	tatgtcattg	2280
gatttgttgt	gccaaatcaa	aaggaaacta	ctgaactagc	tcgaaagaaa	ggacttaaa	2340
ggacttggga	ggagctgtgt	aacagttgtg	aaatggaaaa	tgagctactt	aaagtgtctt	2400
ccgaagctgc	tatttcagca	agtctggaaa	agtttgaaat	tctagtataa	attcgtttga	2460
gtcctgaacc	gtggaccctt	gaaactgggt	tagtgacaga	tgccctcaag	ctgaaacgca	2520
aagagcttaa	aacacattac	caggcggaca	ttgagcgaat	gtatggaaga	aaataattat	2580
tctcttctgg	catcagtttg	ctacagtggg	ctcagatcaa	ataggaaaaat	acttgaaatg	2640
catgtctcaa	gctgcaaggc	aaactccatt	cctcatatta	aactattact	tctcatgacg	2700
tcaccatttt	taactgacag	gattagtaaa	acattaagac	agcaaaactg	tgtctgtctc	2760
ttctttcttc	ttccccgcca	ccaacttact	ttaccacctt	tgactgtact	tgctagtatg	2820
agaatttttc	tgaatcatat	tggggaagca	gtgaattttt	aaacctcaag	ttttaaacat	2880
gatttatatg	ttctgtataa	tgttcagttt	gtaacttttt	aaaagtttgg	atgtatagag	2940
ggataaatag	gaaatataag	aattgggttat	ttggggggtt	ttttacttac	tgtattttaa	3000
aatacaaggg	tattgatatg	aaattatgta	aattttcaat	gcttatgaat	caaatcattg	3060
ttgaacaaaa	gatttgttgc	tgtgttaatta	ttgtcttgta	tgcatattgag	agaaataaat	3120
ataccataac	ttatgtttta	agaagttgag	atcttgtgaa	tatatgcctg	tcagtgtctt	3180
ctttatatat	ttatttttta	ttagaaaaaa	tgaagtttgg	ttgggtgatg	atgaaacaaa	3240
atagcaagag	agggttatag	tttaatatga	agggagataa	cacagcatgt	gtagcaccag	3300
ttgataattg	gtctctagta	gcttactgtc	aaaatgttca	atgaagtctt	ctgttcatct	3360
gttgaaacta	ggaaaatacc	caaacttaaa	tggagaagat	ctgaaagaga	ggatagaatt	3420
taaagaacaa	gagtataata	agttattctt	tgaatttttc	attgactata	tgtacattga	3480
gttatctata	tttgtaaaaa	aaattagtcat	ggaaaattat	tctatcccaa	agtcctcttt	3540
tagtctagat	aatcattatt	tcatttttaa	attagtgttt	ttcatagtgt	gcactgatgc	3600
gtgtatggat	gtgtgtgagt	cagtggtagc	ttattttaaa	agcaccttat	cctttctccc	3660
ataacctttg	tacactaaaa	aatgaaagaa	tttagaatgt	atgtgatgat	agcattctca	3720
ctaagacaca	tgagaattta	actttataac	cgcgtgagtt	aagatttaat	tcatagggtt	3780
tgatgtcatt	gttgaagtta	tttgtaattc	agaaaccttg	cttgtgtgat	acatagtaag	3840
tctcttcatt	tattactgct	tgtctgttgt	tatatctgga	ttatcaaaag	caatagtgca	3900
ccaattaaga	tgtgtctcaa	tcaggactta	aatcataggc	accacatttt	tcatgtcaga	3960

WO 00/26350

PCT/US99/25820

ctagttactt tggtgattct cagttactgt aggcatacaa aggcacaaaat caaaaaaaaaa 4020  
aaaaagg 4027

<210> 10  
<211> 1808  
<212> DNA  
<213> Homo sapiens

<220>  
<221> unsure  
<222> 1597, 1650, 1656, 1713, 1719, 1754, 1778, 1791, 1794, 1796,  
<221> unsure  
<222> 1800  
<223> a or g or c or t, unknown, or other

<220>  
<221> misc\_feature  
<223> Incyte ID No: 3324214CB1

<400> 10  
ctgtcaggaa ggaccatctg aaggctgcaa tttgttctta gggaggcagg tgctggcctg 60  
gcctggatct tccaccatgt tcctgttgct gacttttgat agcctgattg tcaaccttct 120  
gggcatctcc ctgactgtcc tcttcacct ccttctcggt ttcacatag tgccagccat 180  
ttttggagtc tcctttggta tccgcaaaact ctacatgaaa agtctgttaa aaatctttgc 240  
gtgggctacc ttgagaatgg agcgaggagc caaggagaag aaccaccagc tttaacagcc 300  
ctacaccaac ggaatcattg caaaggatcc cacttacta gaagaagaga tcaaagagat 360  
tcgtcgaagt ggtagtagta aggtctctgga caacactcca gagttcgagc tctctgacat 420  
tttctacttt tgccggaaag gaatggagac cattatggat gatgagggtga caaagagatt 480  
ctcagcagaa gaactggagt cctggaacct gctgagcaga accaattata acttccagta 540  
catcagcctt cggctcacgg tcctgtgggg gttaggagtg ctgattcggt actgctttct 600  
gctgccgctc aggatagcac tggctttcac agggattagc cttctgggtg tgggcacac 660  
tgtgtgggga tacttgccaa atgggaggtt taaggagttc atgagtaaac atgttcaact 720  
aatgtgttac cggatctgct tgcgagcgt gacagccatc atcacctacc atgacagga 780  
aaacagacca agaaatggtg gcatctgtgt ggccaatcat acctcacga tcgatgtgat 840  
catcttgccc agcgatggct attatgccat ggtgggtcaa gtgcacgggg gactcatggg 900  
tgtgattcag agagccatgg tgaaggcctg cccacacgtc tggtttgagc gctcggaagt 960  
gaaggatcgc cacctgggtg ctaagagact gactgaacat gtgcaagata aaagcaagat 1020  
gcctatcctc atcttcccag aaggaacctg catcaataat acatcgggtg tgatgttcaa 1080  
aaaggggaagt tttgaaattg gagccacagt ttacctgtt gctatcaagt atgacctca 1140  
atttgccgat gccttctgga acagcagcaa atacgggatg gtgacgtacc tgctgcgaat 1200  
gatgaccagc tgggccattg tctgcagcgt gtgttacctg cctcccatga ctgagagagg 1260  
agatgaagat gctgtccagt ttgcgaatag ggtgaaatct gccattgcca ggcagggagg 1320  
acttgtggac ctgctgtggg atggggggcct gaagagggag aaggtgaagg acacgttcaa 1380  
ggaggagcag cagaagctgt acagcaagat gatcgtgggg aaccacaagg acaggagccg 1440  
ctcctgagcc tgcctccagc tggctggggc caccgtgcgg ggtgccacg ggctcagagc 1500  
tggagtggcc gccgcgccc ccactgctgt gtcctttcca gactccaggg ctcccagggc 1560  
tgctctggat cccagggtt cggctttggc gagccgnagg ggatcctgtg gaccggcgcc 1620  
acttaccttg gtggtttaaa cggatgtgtn ggtttngacc aggagggatg cttgtttttt 1680  
taaatagtgt ttgaggatgc attaatgaa ttnaatttna agttttgggt tatgttggga 1740  
attggcatgt tttnttgggt ggtaagggtt ttaagcctt tcagggtttt ngngnaatn 1800  
tttaggat 1808